DOI:
10.1037/edu0000436

Document Version
Peer reviewed version

Citation for published version (APA):
Tenenbaum, H., Winstone, N., Avery, R., & Leman, P. J. (2019). How Effective is Peer Interaction in Facilitating Learning? A Meta-Analysis. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000436

Abstract

Decades of research indicate that peer interaction among children and adolescents can be beneficial for learning and development. Less, however, is known about which features of interaction may be effective in promoting learning. This meta-analysis examined results from 62 articles with 71 studies into peer interaction, involving a total of 7,105 participants aged 4 to 18 years. The meta-analysis found that peer interaction was effective in promoting learning in comparison with other types of learning conditions. Moderator analyses suggested that learning from interaction with peers was as effective as learning from adults one-on-one, and more effective than children learning individually. Peer interaction is also more effective when children are specifically instructed to reach consensus. Findings point to theoretical considerations for developmental work and practical implications for the effective use of peer interaction techniques in the classroom.

How Effective is Peer Interaction? A Meta-Analysis

From science to the arts and across a range of age groups, individuals often learn through interaction with peers. In children, peer interaction can facilitate cognitive development (Topping, Buchs, Duran, & Van Keer, 2017). However, the dynamics of interaction between children who are engaged in learning activities is often not systematically managed or planned in the classroom (Baines, Blatchford, & Chowne, 2007). Moreover, peer interaction is not always effective. Therefore, knowing more about the factors that contribute to successful peer interaction is essential both from a theoretical perspective and to inform effective classroom practice. In this paper, we present a meta-analysis of the effectiveness of peer interaction across a variety of domains, exploring moderators of the effectiveness to inform theory and practice in this area.

Developmental Theories: Peer Interaction and Intersubjectivity

Peer interaction is typically conceptualised as stemming from constructivist theories of learning, such as the radical constructivism of Jean Piaget, and the social constructivism of Lev Vygotsky. While Vygotskian approaches are typically considered more socially grounded, both theories and the subsequent research inspired by them place peer interaction at the heart of many developmental processes (Tudge & Rogoff, 1999). In both, the notion of intersubjectivity, that is, shared meaning between interactional partners, is key to explaining possible learning benefits from peer interaction. Piaget (e.g., Piaget, 1932, 1959) viewed peer interaction, distinct from adult-child interaction, as an important means of promoting intersubjectivity and subsequent social development. For Piaget, the child is an active participant in constructing knowledge. According to this approach, adult-child interaction is less effective in promoting learning because the adult's natural authority leads to asymmetry which renders the child a more passive recipient of knowledge and instruction (Davis & Winstone, 2017). With peers, in contrast, children are better able to question others as free and active participants in social discourse, argument, and learning (Castellaro & Roselli, 2014). It is this active engagement from all interlocutors that leads to learning gain. Studies across a range of domains, from both Piagetian (e.g., Dimant & Bearison, 1991; Druyan, 2001; Golbeck & Sinagra, 2000; Kruger, 1992; Light & Littleton, 1994; Slavin, 1992) and from Vygotskian (e.g., Garton & Pratt, 2001; Samaha & DeLisi, 2000; Tudge, Winterhoff & Hogan, 1996; Underwood, Underwood & Wood, 2000) frames have demonstrated benefits for peer interaction over and above individual or independent learning.

Research into peer interaction and social learning, in line with Piaget's theory, has sought to demonstrate the benefits of conversation between peers on developmental processes. For instance, Doise, Mugny, and Perret-Clermont (1975) demonstrated the benefits of peer interaction over independent learning for children's grasp of concepts of conservation. Doise and Mugny (1984) have argued that peer interaction is beneficial because it generates socio-cognitive conflict (akin to cognitive conflict) which prompts children to appreciate and consider another child's different perspective and, in turn, develop by adjusting their understanding of a situation accordingly. Subsequent research has sought to explore how far social processes and aspects of the child's social identity can promote or inhibit development (e.g., Psaltis, Duveen, & Perret-Clermont, 2009) or the ways in which children at different ages respond to the social dynamics of peer interaction (Leman, 2015).

Similarly, Vygotsky (1978) emphasised the role of intersubjectivity in development. Subsequent approaches, developed out of Vygotsky's theories, have focussed on development as a process imbued with social and cultural context. While the socio-cultural context remains vital to conceptualise learning and development, Vygotsky proposed a zone of proximal development (e.g., Nyikos & Hasmioto, 1997) where interactions between children (interpsychological) are appropriated into cognitive and metacognitive processes (intrapsychological). Thus, children learn from peers with relative expertise through sharing of knowledge, and a mutual understanding that is generated in social interaction (Jaworski, 1994).

This form of interaction is central to programmes such as cooperative learning (e.g., PALs) and peer tutoring, which involve structured forms of interaction. In PALs, for example, two pupils with different levels of achievement work together while the roles of tutor and tutee are alternated. Children are additionally taught to work in groups and receive targeted subject lessons from teachers, such as summarizing paragraphs and making predictions (Fuchs, Fuchs, & Burish, 2000; Fuchs et al., 2002; Fuchs, Fuchs, Kazdan, & Allen, 1999). PAL programmes have been found to produce positive learning outcomes, particularly with younger children, and low-income and minority urban children (Rohrbeck, Ginsburg-Block, Fantuzzo, & Miller, 2003). Peer tutoring programmes are another type of structured programme found to help children learn. In peer tutoring programmes, one child, who is typically older or more advanced cognitively, is trained to work with another child. Across 72 studies of peer tutoring, Leung (2015) found that peer tutoring was effective, with a medium effect size (Hedge's g = .59).

One concern is that such programmes require significant time and resource investment from teachers, and differ from the more common naturally-occurring interaction that takes place in everyday learning contexts. In typical classrooms, children spontaneously engage in interaction while completing learning tasks, and teachers often directly encourage children to engage in conversation. It is the effectiveness of these conversations that take place between untrained peers that forms the focus of our analysis.

Exploring the efficacy of peer interaction

A key feature of our focus is that it is characterised by spontaneous, bi-directional interaction, involving "the use of small groups of students working together to achieve common goals of learning" (Topping et al., 2017, p. 5). This stands in contrast to situations where children are trained to learn through interaction (i.e. cooperative learning; see Roseth, Johnson, & Johnson, 2008), or where one peer tutors another (i.e. peer tutoring; see Leung, 2015).

We chose to focus specifically on these more spontaneous forms of interactive inquiry learning because this reflects the kind of learning that might be influenced by peer interaction in typical classrooms. From this perspective, a meta-analysis is important on both theoretical and practical grounds. On a theoretical level, such an approach will isolate the effects of peer interaction in typical learning situations from environments where children are trained prior to the peer conversation to engage in interaction. Indeed, understanding the effectiveness of untrained peer conversations compared to child-adult conversations informs us about the degree to which Piagetian versus Vygotskian theoretical perspectives capture children's learning. In particular, in this approach we are able to explore the cognitive mechanisms underpinning learning through peer interaction. From a Piagetian view, peer interaction facilitates learning because intersubjectivity creates socio-cognitive conflict. If it is the subsequent restoration of equilibrium that is fundamental to this process, then we would expect greater learning where the task specifically requires children to reach consensus through interaction. This approach also leads to very clear practical implications; as unstructured peer interaction is more common in everyday pedagogic practice, understanding the moderators of its effectiveness can inform teachers how best to facilitate learning through peer interaction, without engaging in more resource-intensive schemes that involve training and further input.

Previous narrative reviews of collaborative learning identify factors that may act as moderators of the extent to which learning gains result from peer interaction (e.g., Davis & Winstone, 2017; Sills, Rowse, & Emerson, 2016). Such factors include the age of the children involved in interaction, the gender mix of the group, group size, and the area of learning. We also examined characteristics of the studies themselves such as assessment pattern (i.e., whether the tasks were assessed at pre- and post-test or post-test only), the time (or delay) in assessing learning gains, whether consensus was required, the comparison (i.e., children working alone or with adults). These factors fall into three broad categories of moderators: study design factors (assessment pattern, post-test delay, consensus required, and comparison measures); interactional dynamics factors (age, gender, group size); and area of learning.

Study Design Factors

The potential moderating effects of these factors can inform both theoretical and practical understanding of the impact of peer interaction, and under which circumstances these effects are most reliably demonstrated.

Assessment pattern. Contrasting theoretical approaches, alongside practical considerations about how studies can be designed for developing classroom interventions, have also led to differing approaches to the design of studies into peer interaction. Classic Genevan studies from a Piagetian perspective typically used a classic pre-test and post-test design (e.g., Doise & Mugny, 1984), often outside of a traditional classroom learning situation. In contrast, some studies, especially in more applied settings, may include only learning or developmental outcomes for children at post-test (i.e., without a measure of knowledge and ability prior to interaction). Given that pretest assignment controls for differences between groups, we expected greater change from peer interaction compared to the control when pretest measures were taken than when they were not.

Post-test delay. A further aspect of variation in studies of peer interaction concerns when learning gains are assessed. In many classic experimental designs, post-test measures are taken immediately after peer interactions (Messer, Joiner, Loveridge, Light, & Littleton, 1993). Often, differences between peers working together and individuals are small when tested immediately after an interaction (e.g., To, 2015) or do not favour the peer group (e.g., Messer et al. 1993). Other times, assessments may be taken from when peers work together (Leman & Oldham, 2005). In such cases, better performance again may not be observed in peer interaction compared to individuals working alone. Howe, McWilliam, and Cross (2005) posit that learning gains are greater after a delay because children have time to incubate ideas discussed during peer interaction, which primes further learning. In support of this idea, Howe et al. (2005) assessed children's scientific reasoning after a delay arguing that gains are more apparent after a larger interval of time than immediately after an interaction. Based on Howe et al.'s incubation idea, we posit that learning gains should be greater after a delay than when assessed either during a group task or immediately afterwards.

Requirement for consensus. Studies of peer interaction differ according to whether task instructions require children to reach consensus as part of their dialogue. For example, in the moral reasoning domain, children made greater gains on a moral judgement task where they were required to reach a consensual answer compared to engaging in open-ended discussion (Maitland & Goldman, 1974). Consensus building should lead to greater socio-cognitive conflict and subsequent restoration of equilibrium (Piaget, 1932). We expect that the process of reaching consensus will be more beneficial for children's learning than not being asked to reach consensus.

Adult versus child comparison. To whom the peer condition is compared also merits consideration in understanding how the characteristics of a task may influence learning. For example, Radziszewska and Rogoff (1988) assigned 9- and 10-year-children to complete a planning task with mothers or a peer. Children who worked with mothers were found to learn more efficient planning strategies than those who worked with peers. Vygotskian (1978) perspectives suggest that children learn more from adults than from peers who are of a more similar level. Through one-on-one interaction, adults should well placed to scaffold children's learning (Wood & Middleton, 1975). More frequently, however, the literature on peer interaction compares peers to individual children working on their own (e.g., Blaye, Light, Joiner, & Sheldon, 1991; Russell, 1981; To, 2015). From a Vygotskian perspective, we expected that children would learn more from individual adults than peers.

Same task comparison. Another way in which studies differ is the comparison condition, which can be assessed in two way. First, sometimes the comparison group does an identical task (individually or with an adult) or does not complete a related task and functions as a true control condition. For example, Leman, Skipper, Watling, and Rutland (2016) had children work on science tasks with others or not complete any related tasks. Not surprisingly, children who engaged in the related learning task outperformed children who did not engage in learning related science material. More typically, however, children learning in pairs or groups are compared with children who engaged in a similar activity independently (Almasi, 1995; Asterhan, Schwarz, & Cohen-Eliyahu, 2014; Bertucci, Conte, Johnson, & Johnson, 2010; Leman & Oldham, 2005). When To (2015) compared students working alone on the same task as those working in groups, there was no statistically significant advantage for those working in groups. Thus, we expect the differences to be greater when students study the materials on their own rather than serve as a true control.

Interactional Dynamics

Interactional dynamics are likely to be important because peer interaction involves the coordination of shared activity towards a goal. From a theoretical perspective, the moderating effects of these factors can inform understanding of the impact of the degree of symmetry or asymmetry between interactional partners. From a practical perspective, these factors can inform pedagogic practice involving peer interaction.

Age. The strategies used for collaboration may differ with age shifting (at 9-10 years) from a view of interaction as an opportunity to transmit knowledge, to a view of interaction as meaningful collaboration and co-construction (Leman, 2015). As a result, we would expect age differences in children's ability to collaborate and their ability to learn from such collaborations. More specifically, we would expect older children to learn more from peer interactions than younger children.

Gender. Gender is another factor that might influence the effectiveness of peer collaboration. A significant body of research has explored how far the gender of children engaged in interaction affects learning (e.g., Strough, Berg, & Meegan, 2001). Such research suggests that the conversation and interaction styles of girls and boys differ (Leman, Ahmed, & Ozarow, 2005; Leman & Tenenbaum, 2017) such that girls participate more when not outnumbered by or partnered with boys (Webb, 1989). However, certainly by later childhood, these conversation and interaction effects do not appear to feed through to influence learning outcomes (Leman & Björnberg, 2010). Understanding the role of gender in peer interaction adds an important dimension to developmental theory because it can demonstrate the significance of social identities and social dynamics in peer interaction. In addition, understanding the part played by gender is an important consideration for educators who understand the social dynamics within a classroom and seek to optimise the learning gains of their pupils. Thus, we explored the gender composition of the dyad as a predictor of learning gains.

Group size. Another key aspect of social context is the size of a group engaged in interaction. Pairs of children (i.e., groups of 2) will likely facilitate higher quality and more intensive interaction than larger groups because it is difficult for one child not to be engaged (Webb, 1989). However, more children involved in interaction also increases the possibility of different perspectives being introduced to a conversation. In this respect, more perspectives may bring greater opportunity for the transfer of knowledge or could generate greater socio-cognitive conflict (Fuchs, Fuchs, Kazdan, Karns, Calhoon, Hamlett & Hewlitt, 2000). It may be that there is an optimal group size; for instance, more than 6 children in a group may lead to social loafing (i.e., exerting less effort than optimal, see Topping et al., 2017). We expected that children in larger groups would outperform those in smaller groups.

Area of learning

The area of learning also varies across studies of peer interaction. Research has examined peer interaction in areas including conservation, creativity, moral judgments, general reasoning, scientific reasoning, mathematical reasoning, spatial reasoning, spatial conservation, and memory tasks. Area effects may be attributable to the different learning objectives or complexity or specificity of the subject matter (e.g., Bloom *et al.*, 1956) and thus, be more or less amenable to learning in a social, interactional context.

To answer the question of the degree to which peer interaction is an effective pedagogical method, we conducted a meta-analysis seeking to address the following research aims:

1. Does research support the effectiveness of peer interaction in facilitating learning, in comparison to other types of learning?
2. Is the effectiveness of peer interaction influenced by features of the study design (post-test delay, assessment pattern, instructions for consensus, same task comparison, adult versus child comparison)?
3. Is the effectiveness of peer interaction influenced by interactional dynamics (age, gender, group size)?
4. Does the effectiveness of peer interaction differ according to the area of learning?

Method

Study Sample

Our meta-analysis included 62 separate publications that included 71 separate samples. These studies included a total of 7,105 children between 4 and 18 years of age. Samples sizes ranged from 26 to 390 children in each study. The majority of studies were conducted in the US (*n* =32) and the UK (*n* =23). Studies were also conducted in Australia, Canada, Chile, Cyprus, Germany, Israel, the Netherlands, Portugal, Singapore, Sweden, Switzerland, and Russia. Most studies (*n* = 49) provided no information about ethnicity, which may be a consequence of the date of publication. Although some included Asian and Asian-American, African-American, Latino/a, and middle-eastern young people, the majority of the samples that reported information about ethnicity were ethnically homogeneous samples of White European and European-American children.

Literature Search

We identified reports examining peer interaction through a variety of sources. We included reports from as early as we could find up to and including 2017. We did not have any geographical restrictions. All articles were published in English. The earliest article identified was Hudgkins (1960). First, we looked for articles, book chapters, and dissertations using computerized searches of *PsychINFO, ERIC,* and *Dissertations Abstracts International.* Literature searches were conducted using different keywords (i.e., peer, peer learning, collaborative learning, cooperative learning, peer interaction, peer conversation) alone and in combination with other words (i.e., children, adolescents, young people), which yielded 11,570 results. Second, we compiled all the citations from Davis and Winstone (2017) and Sills, Rowse, and Emerson (2016), with these sources representing relevant syntheses on this topic. Third, we then examined the profile pages of authors of each article included to see if they had conducted other studies that were not included in our database. Fourth, studies were identified from citations in articles and from forward citations in databases. Finally, we examined the conference schedules from the American Education Research Association, the European Association for Developmental Psychology, and the Society for Research in Child Development. The abstracts were scanned for relevance, which left 210 articles in the database. After using additional criteria (see below), 62 articles with 71 separate studies were retained. At least two authors together checked each article before exclusion and a reason was noted on a spreadsheet shared between all authors. Figure 1 displays the PRISMA diagram of this procedure.

The selection criterion was that studies had to test directly for differences between working in peers and a comparison group (e.g., individuals alone, individuals with adults, etc.) in learning in children 18 years or younger.

Exclusion criteria precluded the use of several potentially relevant studies. First, studies that reported pre-interaction, individual test scores and then scores after children worked in peers (post-test) were not included if they did not include a comparison group because children could improve on a task over time with no intervention. We excluded 68 studies for not having comparison conditions. Second, the comparison group needed to differ from the peer group only on the peer interaction element. If the peer group engaged in an additional task (e.g., presenting to the class in addition to peer interaction) and the control group did not (e.g., Bunrasi, 2012), we did not include they study because we were unable to isolate the effects of peer interaction from other factors. Because of the focus on peer interaction rather than programmes that come from other theoretical traditions, we did not include studies using PALs (Fuchs et al., 2002), cooperative learning (e.g., Nichols & Miller, 1994), or peer tutoring (Leung, 2015). Third, studies that focused on participants whose mean ages were older than 18 years were excluded. Fourth, if the peer group received additional support different from the control group, we excluded the study. Fifth, studies that did not report quantitative data were excluded. Because we did not want to perform a sign test, which simply records whether the data supports or disconfirms a difference (see Dixon & Mood, 1946), we did not include articles that did not provide useable statistical information. However, before discarding any articles, authors were contacted for information that could be included in the meta-analysis. We wrote to 16 authors. Six replied with information about nine separate articles. Following recommendations by Borenstein, Hedges, Higgins, and Rothstein (2009), if a study reported no significant difference without further statistical information, data from that study was recorded as $d = .00$ and added to the other comparison. Only two comparisons did not report exact p values or statistics (e.g., F-value, means and standard deviations, etc).

Information Extracted

A coding scheme was created based on variables that could serve as moderators of effects. Table 1 describes these moderators.

To start, we considered features of the design. First, we coded whether studies used a pretest-posttest design or a posttest only assessment pattern as part of the design. Second, we examined when the testing of interest occurred. We recorded whether it occurred while groups worked together, immediately after, after a delay of half an hour or more, or whether no information was reported separately for time. Third, we recorded whether instructions to students explicitly asked them to reach agreement, consensus, or not. We included studies in the former category who instructed student to try to arrive at consensus, but if this was not possible, they needed to understand the other person's argument (Kuhn, Shaw, & Felton, 1997; Murphy & Messer, 2000; Schwarz & Linchevski, 2007). Fourth, we recorded whether the comparison group did the same task (a true control). Some studies had 50% or 75% of the sample do the same task as the peer groups (intervention), but did not report this information separately, so we recorded this as mixed. Fifth, we recorded to whom children in the intervention groups were compared (i.e., individual children or adults). For example, Gauvain and Rogoff (1989) compared peer performance to children working alone as well as children working individually with an adult.

Age was the first moderator related to the interactional dynamics we considered. We split samples into whether they included primary-aged students (up to age 10 years) or older (11 to 18 years). Studies were coded as mixed if they spanned both age groups. If an article reported data separately for two different age groups independently (e.g., 6-year-olds and 12-year-olds), we counted these effects as separate samples. The youngest age group included in the studies was age 4 years. Second, we recorded whether the groups were composed of single-gender, mixed-gender, both mixed- and same-gender groups, or not reported. If an article reported statistics separately for samples by gender of pairs (e.g., Samaha & DeLisi, 2000), we entered that information separately for the moderator analyses. We had originally aimed to look at gender of the participants, but only five studies conducted single-gender samples so this was insufficiently powered. Third, we examined the size of the peer group by comparing studies that assigned children to dyads versus larger groups and compared them to a control group (e.g., Roazzi & Bryant, 1998).

Finally, we looked at the area of learning under investigation, which included general reasoning tasks that did not involve subject domains or content (e.g., puzzle completion, deciding when to cross a road, sorting blocks by color and shape, etc), conservation (e.g., liquid, mass), creativity (e.g., Torrance tests of creativity), memory (for words), moral reasoning, spatial planning and reasoning (e.g., route planning), and spatial conservation tasks (e.g., three mountains problem; Piaget & Inhelder, 1956). We also created two other categories based on subject knowledge (see Alexander, Jetton, & Kulikowich,1995, for an explanation of domain versus subject knowledge) for science and mathematics which included scientific reasoning (e.g., reasoning about topics within the scientific domain that involved scientific content, such as evolutionary theory, buoyancy, photosynthesis), mathematical procedures and reasoning (e.g., arithmetic, word problems). Although we had tried to split the mathematical domain into reasoning versus procedural knowledge, studies using word problems rely on arithmetical as well as reasoning knowledge (e.g. Hudgins, 1960) and thus, were coded as mathematical reasoning. Similarly, studies of science learning, such as Chan (2001) tested students' evolutionary knowledge, which relied on content knowledge as well as understanding of processes. Thus, such studies were coded as scientific reasoning. Finally, we recorded to whom the children in the peer interaction condition were compared.

Reliability

Coding for moderators was accomplished with recommendations from the four authors who decided on moderator codes to include the range of conditions. The first two authors separately coded 16 articles for all the moderators except area of study. Area of study and requirement for consensus was coded by one of the authors with a trained research assistant. Reliability on each of the moderators and statistics across the article was found to be consistently high leading to kappas ranging from .73 (e.g., on whether the design was pretest posttest or pottest only) to 1.00 (e.g., size of the group, area of study, consensus), which indicates perfect agreement. We also discussed all articles with unclear statistical information. All disagreements were resolved through a discussion of how best to classify the variable in question both within the context of the study and the purposes of analysis. The first author then completed all coding.

Effect Size Calculation

We used CMA Version 3 (Borenstein, Hedges, Higgins, & Rothstein, 2011) to calculate effect sizes. We used a random effects model because there was variability in the design and location of studies conducted. Effect sizes were coded so that a negative effect size indicates that participants in the compared instructional conditions evidenced greater learning than participants in peer groups, whereas a positive effect size indicates that participants in the peer conditions evidenced greater learning than participants in the compared instructional conditions. The majority of studies reported means and standard deviations for each group.

Unit of Analysis. As the unit of analysis, group samples of studies and comparisons were considered separately. *Studies* as a unit of analysis referred to individual experiments with different participants. Studies, thus, treat multiple experiments reported within a single article as separate studies if they involved different participants or two age groups of participants reported separately within a single article. Three articles contributed two separate experiments and three contributed two separate studies. Two articles reported age groups separately that fit into the ages examined separately in this meta-analysis. *Comparisons* were also used as a unit of analysis in two analyses in a separate dataset than the one used to examine studies. These three analyses (comparison group, same task, gender) were of interest to the present meta-analysis. Analysis at the level of comparisons refers to counting each individual statistical comparison as an independent contribution. Although multiple comparisons reported for a single sample violate assumptions of independence, analysis at this level was required to test for effects of these two moderating variables. Six studies had different comparison groups in which children in peer groups were either compared to individuals working alone or individuals working with adults (Druyan, 2001; Gauvain & Rogoff, 1989; Howe et al., 2013; Howe et al., 2016; Murphy & Messer, 2000; Tolmie et al., 2005). The latter group consisted of different children. In this case, we divided the number of children in the intervention condition by the number of comparison groups so that studies with more than one comparison would not weight the analysis disproportionate to its sample size following procedures set out by Borenstein, et al. (2011). This data set was also used to examine whether children completed the same task or not because these studies used a control group in which children were tested twice without completing a task. However, two studies were not included in these analyses (Howe, McWilliam, and Cross, 2005 study 2 and study 3) because comparisons were not broken down by whether children completed the same task on their own or not. To examine instructions for consensus, Maitland and Goldman (1974) contributed two studies because the instructions were varied with one condition asking participants to arrive at consensus, whereas one did not. In this case, we divided the control group in half for each comparison so that each participant only contributed one data point. For gender analyses, one study used two different gender groups (groups of boys, groups of girls, mixed-gender groups) with different children in each peer interaction group (Samaha & DeLisi, 2000). In this case, we divided the number of children in the comparison group by three so that they would not have a disproportionate weight.

Because some studies reported scores from each individual in the peer condition and others took a mean from the peer group or selected one child from the group to represent the group score, studies using the first strategy would have been weighted unequal to the other types of studies. After discussion with Michael Borenstein, we weighted studies by the number of participants in each study. Thus, if a study used one child's score for a pair of children (e.g., Tudge, 1989), we used the mean and standard deviation for this group and counted the actual number of participants rather than the number reported by the authors.

Hedge's g values are reported here as calculated by the CMA program. *Hedge's g*s between .20 and .50 indicate a small effect size, between .50 and .80 indicate a medium effect, and greater than .80 indicate a large effect (Hedges, 1981). Of course, the effect size alone does not determine statistical significance and we determined the statistical significance of effect sizes based on the *p*-values of the resultant *Z*-scores.

When multiple effect sizes were reported, they were averaged to form a single effect size using the formula recommended by Borenstein et al. (2011):

$$\bar{Y} = \frac{1}{m}\left(\sum_{j}^{m} Yj\right)$$

(*Equation 1)*

In this formula, m signifies the number of tests in a study that we combined, Y are the values of the different effect sizes in a study, and j indicates the particular study (see Borenstein et al., 2011, equation 24.4).

Post-hoc Tests

After grouping the effect sizes by a particular moderator and finding significant heterogeneity among different levels of the same moderator, each level was compared to all others within the CMA program, indicated by *Q*, to determine if the effect sizes between the groups were significantly different from one another. Post hoc *p*-values were adjusted for the number of comparisons conducted using Bonferroni corrections. For example, post hoc comparisons of the category, gender composition of peers (e.g., same-gender, mixed- and same-gender, mixed-gender), required three comparisons and consequently led to a set alpha level of 0.02 (.05 divided by 3) for levels to be considered significantly different from one another.

Results

Does research support the effectiveness of peer interaction in facilitating learning in comparison to other types of learning?

A total of 71 samples from 62 articles compared peer interaction to other types of learning. Using a random-effects model, studies computed separately for group samples had a mean weighted effect size of Hedges' $g = 0.40$, 95% CI [0.27, 0.54], $p < .0001$. This effect constitutes a small but meaningful effect size. The positive sign indicates that children assigned to peer interaction conditions evidenced greater learning than those assigned to other conditions.

The fail-safe N, which is the number of missing studies with a mean Hedges' *g* of 0 that would make the effect size no longer statistically significant, was 2,669 for the data set using studies as the unit of analysis. We also visually inspected the funnel plot (see Figure 2) to look for whether studies cluster around the effect mean effect size, which would suggest publication bias (Borenstein et al, 2009). We found convergence toward the mean in big studies, which would suggest minimal publication bias (Borenstein et al., 2009). However, there were quite a few outliers. We also conducted correlation coefficients using Spearman's rho in SPSS. Sample size was not normally distributed, which is why we used non-parametric statistics. There was not a statistically significant relation between sample size and Hedges' *g* effect sizes, r (69) = -.21, p = .07.

We also examined impact of publication bias on the size of the effect reported. We included three unpublished studies because research shows that studies with null effects are often not published, which is known as the file-drawer problem (Cook & Therrien, 2017; Polanin, Tanner-Smith, & Hennessy, 2016; Rosenthal, 1979). As a result, a common criticism of meta-analysis is that meta-analysis miscalculates the overall effect size by not including unpublished studies (Esterhuizen, & Thabane, 2016). To combat this issue, experts on meta-analysis recommend including unpublished studies (Cook & Therrien, 2017; Esterhuizen, & Thabane, 2016; Polanin, Tanner-Smith, & Hennessy, 2016; Rosenthal, 1979). We compared published (Hedge's g = .38, 95% CI [.25, .50] with unpublished contributions (Hedge's g = .95, 95% CI [-.42, 2.32]). There was no statistically significant difference in effect sizes, Q (1) = .66, p = .42.

To further assess publication bias, we used Duval and Tweedie's (2000) Trim and Fill technique. The Trim and Fill technique removes the most extreme small studies from either the negative or positive side of the population mean. Data are trimmed until a symmetrical funnel plot is produced, which computes a new effect size based on the mean of the re-computed symmetrical funnel. At this point, the trimmed studies are filled back into the plot and a mirror image created so that the plot remains symmetrical (Borenstein et al., 2011). If the effect size change is small, once the plot has been trimmed and imagined studies are accounted for, then the impact of publication bias in the sample is considered to be low. When we trimmed the studies from the left of the mean using a random effects model, there was no change in the effect size. The point estimate was 0.40, 95% CI [0.27, 0.54]. When we used a random effects model and trimmed from the right side of the distribution, the imputed point estimate was 0.42, 95% CI [0.29, 0.55]. Together, these indicators suggest that strong evidence of bias does not result from missing studies.

We conducted sensitivity analyses by removing each study and re-calculating the statistics. The Hedges' *g*s ranged from .37, 95% CI [.25, .50] to .42, 95% CI [.29, .55] when we did this. When the study with the largest number of participants was removed (Asterhan et al., 2007), the Hedge's g was .41, 95% CI [.28, .55]. In sum, these indicators suggest minimal bias in the findings.

Moderators

Homogeneity analyses for the group sample indicated that the effects were highly heterogeneous, Q_w (70) = 344.23, $p < .0001$. The large value suggests that variability in results may not be due to sampling error alone given the size of the sample (Rosenthal, 1991). For this reason, we were able to investigate different moderators.

An advantage of quantitative meta-analytic techniques is the ability to examine potential moderators of relations with ample statistical power. In the present meta-analyses, the following potential moderators were investigated: features of the design (i.e., post-test delay, assessment pattern, instructions for consensus, same task comparison, and adult versus child comparison), interactional dynamics (i.e., age of participants, gender of pairs, group size), and the area of learning. Table 3 shows effect sizes, which comparisons were significantly different from each other, and whether the studies in each moderator category demonstrated a significant effect.

Is the effectiveness of peer interaction influenced by features of the study design? The first moderators examined aspects of the design. First, we looked at the assessment pattern of the study and found that whether the study used a pretest-posttest or posttest only did not influence the findings, Q (1) = .02, p = .89, which was in contrast to our hypothesis. Second, in contrast to our hypothesis that delays would increase learning gains in peer interactions, the time of the post-test (during, after, or with a delay) did not moderate the effect, Q (2) = 4.85, p = .18. Third, confirming our hypothesis, when asked to achieve consensus, children in peer groups made greater gains than when no instructions were provided, Q (1) = 11.74, p = .0001. Fourth, disconfirming our hypothesis, the performance or outcomes of children engaged in peer group learning did not differ according to whether children in the comparison conditions did the same task or a different task, Q (1) = .80, p = .37. Finally, supporting our hypothesis, children working individually with adults outperformed peer groups working together, Q (1) = 12.40, p < .0001. Note that peers working together did not differ from children working individually with an adult. There was support for two of our hypotheses.

Is the effectiveness of peer interaction influenced by interactional dynamics? The moderators related to social factors investigated participant characteristics. In contrast to our hypothesis, age, Q (1) = 2.73, p = .10, did not moderate the effect. There was also no effect of gender composition when we compared children in mixed versus same-gender pairs, Q (1) = 1.82, p = .18. Finally, children learn the same amount in groups of two children and in larger groups, Q (1) = 1.88, p = .17. Thus, none of our hypotheses related to interactional dynamics were confirmed.

Does the effectiveness of peer interaction differ according to the area of learning? Finally, the area of interest did not significantly moderate the findings, Q (8) = 11.05, p = .20.

Discussion

The Effectiveness of Peer Interaction

The present study compared findings from 62 papers including 71 separate samples and 7,105 children. The outcomes of this meta-analysis confirm that peer interaction is associated with beneficial learning outcomes compared with other learning conditions, with a small to medium effect size. This is consistent with meta-analyses of other types of peer learning (e.g., Johnson et al., 1981). There is no evidence of publication bias in these results (in terms of sample size or published versus unpublished reports). However, the conclusion that peer interaction is an effective means of promoting academic advancement is qualified by a large amount of heterogeneity in terms of such effects. In other words, although peer interaction facilitates learning, the conditions and means by which that happens varies greatly and depends on a number of moderating factors. The analysis of the impact of various moderators offers evidence about the source of this variation which has both theoretical and practical implications.

Interactional Dynamics

None of the findings on interactional dynamics supported our hypothesis. First, we found no evidence that the age of children engaged in peer interaction exerted an influence on its effectiveness. This indicates that children of any age may benefit from peer interaction; however, this does not mean that they benefit for the same reasons. Previous work has suggested that the strategies used for collaboration may differ with age, shifting from a representation of interaction as an opportunity to acquire knowledge that is transmitted from another, to interaction as a meaningful collaboration (Leman, 2015).

With greater age, however, children may engage in social loafing. Indeed, children younger than third grade (aged 8 to 9 years) are less likely than older students to engage in social loafing (Karau & Williams, 1993), in which less effort extended in group than individual activities. Part of the reason that younger children may engage in less social loafing is because their understanding of others' mental states is less well developed than it is in older individuals (Thompson & Thornton, 2014). Thus, the underlying reasons for why peer interaction supports children at different ages may be complex.

Gender is another social factor that previous research has examined (e.g., Leman & Björnberg, 2010). Gender influences children's conversations from preschool (Leaper & Smith, 2004; Tenenbaum, Ford, & Alkhedairy, 2011), but the present analysis indicates that the gender composition of groups does not feed through to affect learning outcomes. This confirms findings from previous studies indicating that while gendered conversation dynamics may be very much a feature of children's interactions, children can work around these in solving a problem and learning from the interaction (see Leman *et al.*, 2005). Once again, however, it would be hasty to conclude that gender is not an active consideration in classroom group work because each interaction has a social dimension as well as a learning outcome. Moreover, the present analysis did not consider whether different forms of knowledge (e.g., conceptual versus procedural scientific knowledge) may suit boys' interactions better than girls' interactions (Leman, Skipper, Watling, & Rutland, 2016).

The third social factor we examined was group size. Previous meta-analyses identified that smaller group size was associated with higher effectiveness of cooperative peer learning (Johnson et al., 1981). The present analysis found no evidence for group size effects on the effectiveness of peer interaction. This has important implications for classroom contexts where children are often assigned to work in groups of varying sizes for pragmatic reasons. The optimal sized group for an activity may not be a simple thing to prescribe and in the classroom effective learning through peer interaction could require a matching of social dynamics considering features of the task itself and the age of the children involved.

Area of learning

In the present meta-analysis, we included studies using a wide variety of tasks from many areas of learning. Whilst there was no statistically significant effect of this moderator, Hedge's *g* ranged from -.61 for Creativity through to .98 for Conservation. This large effect for conservation tasks is consistent with a significant strand of work in European psychology (e.g., Doise, Mugny & Perret-Clermont, 1975) that sought to unpick the learning benefits of social interaction between peers. These tasks rely on a set of fundamental cognitive processes that can readily trigger socio-cognitive conflict. Typically, conservation tasks set up disequilibrium by pairing a conserver with a non-conserver (e.g., Botvin & Murray, 1975; Weinstein & Bearison, 1985), and many Piagetian and neo-Piagetian tasks rely on qualitative shifts in understanding. These conditions provide scope for alternative representational systems (understanding) to come into conflict in discussion and debate, which can be a catalyst for learning.

Study Design

Aspects of study design including whether children involved in peer interaction did the same or a comparison task, whether the study used a pretest-posttest or merely post-test design, and the timing of the posttest (specifically, whether there was a significant delay after interaction) did not moderate the effect of peer interaction.

However, one element of study design which significantly moderated the effectiveness of peer interaction was the requirement for peers to reach consensus in putting forward a common answer to a task or problem. In tasks where peers were not explicitly instructed to reach consensus, the effect size was small (*g* = .17). In contrast, when consensus was part of the task instructions, the effect size was of medium magnitude (*g* =.61). This finding is in accord with arguments that cognitive conflict is necessary but not sufficient as a catalyst for learning. The key element seems to be the verbal interactions that facilitate negotiation of diverse viewpoints, testing ideas, and seeing things from a different point of view (Forman & Kraker, 1985). In this sense, the cognitive processes underpinning consensus generate social (interpersonal) disequilibrium (Damon & Killen, 1982). However, previous research suggests that consensus alone is not enough to generate learning gains but requires also some further appraisal of consensual positions (Howe & Tolmie, 2003). In this respect, consensus that is achieved as a social process (e.g., through processes of conformity) is less adequate from a perspective of learning and development than consensus achieved through socio-cognitive processes. Involvement in a discussion requiring consensus also facilitates children's emotional and intellectual engagement and investment in the dialogue. Passivity is difficult in a task requiring consensus. Furthermore, social loafing is less likely when consensus is required; the higher cognitive demands have the potential to ensure that social dynamics in the interaction do not lead to domination of discussion and decision-making.

The meta-analysis further reveals that the nature of the comparison, namely whether peer interaction was compared against children working individually or children working with adults, had a moderating effect. Peer interaction only had positive effects when the comparison group was children working individually (g = .45). In contrast, when compared against children working with adults, peer interaction was not positive in its effect (g = -.30). This finding supports the Vygotskian perspective that adults working one-on-one with children can support children's learning through careful scaffolding of children to help them reconstruct knowledge (Vygotsky, 1978; Wood & Middleton, 1975). The processes through which children learn from peers versus adults might differ, but the findings suggest that children need to be actively engaged. Moreover, support is found for both Vygotskian and

Piaget perspectives in this meta-analysis, which we return to later on.

Limitations

One of the most common limitations in meta-analysis is the amount of information included in the primary literature. As a result of lack of information in many studies we could not code for the duration of activities. The vast majority of studies involved a single interaction so we also could not examine the frequency in which peers engaged in the activities. Nor could we examine whether groups composed of girls or boys made greater gains because there were few studies reporting these effects separately, looked at the effectiveness of all-girl or all-boy groups (e.g., Tudge, Winterhoff, & Hogan, 1996). Few studies explicitly stated characteristics of individual children, such as whether they were at risk or had developmental disabilities. Given the lack of information, we would assume that most studies relied on populations of children who were not at risk. All of these limitations suggest arenas for future primary research. For example, whether peer interaction remains more effective over repeated interactions than individuals working alone merits consideration.

Implications for Theory and Practice

These findings are broadly supportive of theoretical approaches that have proposed that social interaction plays an important role in learning (e.g., Piaget, 1932; Vygotsky, 1978), and provide further evidence that conversation and language play a primary role in supporting children's learning as others have argued (deRosnay & Hughes, 2006; Tenenbaum, Leman, Aznar, & To, 2016). More specifically, our key finding that peer interaction is more effective where children are instructed to reach consensus points to the importance of intersubjectivity, socio-cognitive conflict, and restoration of equilibrium as key cognitive processes underpinning the learning gains from peer interaction. Negotiating differences of opinion in the process of reaching consensus creates disequilibrium, requiring children to “try out new ideas” (Palincsar, 1998, p. 350). Similarly, Piaget (1985, p. 10) argued that “disequilibrium forces the subject to go beyond his current state and strike out in new directions”. Vygotsky (1978) also proposed that language can drive developmental processes, where language as a cultural tool facilitates the co-construction of knowledge (Vygotsky, 1981). It appears that scaffolding children’s interaction through the instruction to negotiate a path to learning by working towards consensus is a simple element of task design that can have powerful effects. Whilst Piaget proposed that children working with adults is less effective because the degree of intellectual asymmetry is too great, our findings indicate that working one-on-one with adults is more effective in facilitating children’s learning than peer interaction. In sum, this finding suggests that both Piaget and Vygotsky are correct in suggesting that children can learn from either adults or children as long as children are active members of the conversation.

Of course there is a key pragmatic difference between child-peer and child-adult learning partnerships. Many classrooms have just a few adults to work with a large group of children; the process of children working one-on-one with adults is simply not practical. As a result, children are often asked to work simultaneously and individually on a particular task, which is less effective than peer interaction. That peers can also be effective teachers opens up the possibility for simultaneous learning conversations across an entire classroom of children.

Our analysis supports that peer interaction is important for pedagogic not just practical reasons. Peer interaction appears to be effective across different areas of the curriculum, and in both elementary and secondary education. A key educational implication of our findings is that a simple amendment to task instructions, whereby children working in dyads or groups are simply instructed to reach consensus through interaction, can lead to greater gains in learning. These gains may also extend beyond the focal task. Howe (2015) proposed that the process of exploring and negotiating differences of opinion has the potential to develop metacognitive awareness, which can facilitate future learning in similar situations. Additionally, while peer interaction is associated with effective learning, it should not be forgotten that peer interaction entails many positive social and personal outcomes as well for children including the development of relationships, positive intergroup attitudes, and the rehearsal of communication and collaboration skills. Working with others towards consensus is likely to further facilitate perspective-taking and shared responsibility for task outcomes, which pose further social benefits to children.

Future Research Directions

For researchers, the findings point to the significance of the social and developmental context for learning, and that children's capabilities for effective social interaction inter-relate with their cognitive abilities in often subtle ways. Future research should seek to look more closely at the interplay between social and cognitive factors, as well as considering the influence of broader cultural norms and practices in peer interaction processes (see again, Thanh, Gillies & Renshaw, 2008). Finally, future work should seek to uncover our suggestion that reaching consensus may play a central role in effective peer interaction. Untangling the mechanisms underlying successful peer interaction can contribute to increased learning gains and more effective classroom pedagogy.

Conclusion

Whilst many pedagogic designs require children to work in dyads or groups for practical reasons, key developmental theories point to the potential benefits of this practice for learning on both cognitive and social levels. Palincsar (1998, p.351) argued that "verbal interaction is the key to co-construction and cognitive change". As a cultural tool, language can facilitate learning, enabling children to co-construct knowledge with an interlocutor (Vygotsky, 1981). Furthermore, the act of working with a peer in the context of a learning task has the potential to create disequilibrium through socio-cognitive conflict, with the process of reaching consensus creating the conditions for cognitive growth (Piaget, 1985). Our findings indicate that the benefits of peer interaction can be realised by educators if they create opportunities not just for discussion, but the negotiation of a shared understanding.

References

Alexander, P. A., Jetton, T. L., & Kulikowich, J. M. (1995). Interrelationship of knowledge, interest, and recall: Assessing a model of domain learning. *Journal of Educational Psychology*, *87*, 559-575, DOI: 10.1037/0022-0663.87.4.559.

Almasi, J. F. (1995). The nature of fourth graders' sociocognitive conflicts in peer-led and teacher-led discussions of literature. *Reading Research Quarterly*, *30*, 314-351. doi: 10.2307/747620.

Asterhan, C. S., Schwarz, B. B., & Cohen-Eliyahu, N. (2014). Outcome feedback during collaborative learning: Contingencies between feedback and dyad composition. *Learning and Instruction, 34,* 1-10. doi: 10.1016/j.learninstruc.2014.07.003.

Bertucci, A., Conte, S., Johnson, D. W., & Johnson, R. T. (2010). The impact of size of cooperative group on achievement, social support, and self-esteem. *The Journal of General Psychology: Experimental, Psychological, and Comparative Psychology*, *137*, 256-272. doi: 10.1080/00221309.2010.484448

Bloom, B. S.; Engelhart, M. D.; Furst, E. J.; Hill, W. H.; Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals*. Handbook I: Cognitive domain. New York: David McKay Company.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*: John Wiley & Sons.

Botvin, G. J., & Murray, F. B. (1975). The efficacy of peer modeling and social conflict in the acquisition of conservation. *Child Development*, *46*, 796-799. Stable URL: https://www.jstor.org/stable/1128583

Bunrasi, J. B. T. (2012). *Algebra I Achievement of Eighth Grade Mexican American Students Using Cooperative Learning versus Traditional Instruction.* (Doctoral dissertation). Walden University.

Castellaro, M. A., & Roselli, N. D. (2015). Peer collaboration in childhood according to age, socioeconomic context and task. *European Journal of Psychology of Education*, *30*, 63-80. doi: 10.1007/s10212-014-0228-3.

Cohen, P. A., Kulik, J. A., & Kulik, C-L. C. (1982). Educational outcomes of peer tutoring: A meta-analysis of findings. *American Educational Research Journal, 19*, 237-248. Stable URL: https://www.jstor.org/stable/1162567.

Damon, W., & Killen, M. (1982). Peer interaction and the process of change in children's moral reasoning. *Merrill-Palmer Quarterly, 28*, 347-367. Stable URL: http://www.jstor.org/stable/23086121.

Davis, A., & Winstone, N. (2017). Educational Implications. In A. Slater & G. Bremner (Eds.), *An Introduction to Developmental Psychology* (3rd ed.; pp. 613-644). Chichester, UK: Wiley.

de Rosnay, M., & Hughes, C. (2006). Conversation and theory of mind: Do children talk their way to socio-cognitive understanding?. *British Journal of Developmental Psychology*, *24*, 7-37.

Dimant, R. J., & Bearison, D. J. (1991). Development of formal reasoning during successive peer interactions. *Developmental Psychology*, *27*, 277-284. Doi: 10.1037/0012-1649.27.2.277.

Dixon, W. J., & Mood, A. M. (1946). The statistical sign test. *Journal of the American Statistical Association*, *41*, 557-566. Stable URL: https://www.jstor.org/stable/2280577.

Doise, W., & Mugny, G. (1984). *The Social Development of the Intellect*. Oxford, Pergamon.

Doise, W., Mugny, G., & Perret-Clermont, A. N. (1975). Social interaction and the development of cognitive operations. *European Journal of Social Psychology, 5,* 367-383.

Druyan, S. (2001). A comparison of four types of cognitive conflict and their effect on cognitive development. *International Journal of Behavioral Development, 25*, 226-236. DOI: 10.1080/01650250042000276.

Esterhuizen, T. M., & Thabane, L. (2016). Con: Meta-analysis: some key limitations and potential solutions. *Nephrology Dialysis Transplantation*, *31*, 882-885. DOI: 10.1093/ndt/gfw092.

Forman, E. A., & Kraker, M. J. (1985). The social origins of logic: The contributions of Piaget and Vygotsky. *New Directions for Child and Adolescent Development*, *29*, 23-39.

Fuchs, D., Fuchs, L. S., & Burish, P. (2000). Peer-assisted learning strategies: An evidence-based practice to promote reading achievement. *Learning Disabilities Research & Practice, 15*, 85-91. Doi: 10.1207/sldrp1502_4 .

Fuchs, D., Fuchs, L. S., Mathes, P. G., & Martinez, E. A. (2002). Preliminary evidence on the social standing of students with learning disabilities in PALS and no–PALS classrooms. *Learning Disabilities Research & Practice, 17*, 205-215. Doi: 10.1111/1540-5826.00046.

Fuchs, L. S., Fuchs, D., Kazdan, S., & Allen, S. (1999). Effects of peer-assisted learning strategies in reading with and without training in elaborated help giving. *The Elementary School Journal, 99*, 201-219. Stable URL: https://www.jstor.org/stable/1002313.

Gauvain, M., & Rogoff, B. (1989). Collaborative problem solving and children's planning skills. *Developmental Psychology, 25*, 139-151. Doi: 10.1037//0012-1649.25.1.139.

Gummerum, M., Leman, P. J., & Hollins, T. S. (2013). Children's collaborative recall of shared and unshared information. *British Journal of Developmental Psychology, 31*, 302-317. doi: 10.1111/bjdp.12006.

Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin*, *92*, 490-499. doi: 10.1037/0033-2909.92.2.490.

Howe, C. (2015). *Conceptual structure in childhood and adolescence: The case of everyday physics*. London: Routledge.

Howe, C., & Tolmie, A. (2003). Group work in primary school science: discussions, consensus and guidance from experts. *International Journal of Educational Research, 39,* 51-72. DOI: 10.1016/S0883-0355(03)00073-9.

Howe, C., Taylor Tavares, J., & Devine, A. (2016). Recognition as support for reasoning about horizontal motion: A further resource for school science? *Research in Science & Technological Education, 34,* 273-289. doi:10.1080/02635143.2016.1168393.

Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D., & Skon, L. (1981). Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin*, *89*, 47-62. DOI: 10.1037/0033-2909.89.1.47.

Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology*, *65*, 681- 706. Doi:

Leaper, C., & Smith, T.E. (2004). A meta-analytic review of gender variations in children's talk: Talkativeness, affiliative speech, and assertive speech. *Developmental Psychology, 40,* 993-1027. Doi: 10.1037/0012-1649.40.6.993.

Leman, P. J., & Björnberg, M. (2010). Conversation, development, and gender: A study of changes in children’s concepts of punishment. *Child Development, 81*, 958-971. Doi: 10.1111/j.1467-8624.2010.01445.x

Leman, P. J. (2015). How Do Groups Work? Age Differences in Performance and the Social Outcomes of Peer Collaboration. *Cognitive Science, 39,* 804-820. Doi: 10.1111/cogs.12172.

Leman, P. J., & Oldham, Z. (2005). Do children need to learn to collaborate?: The effect of age and age differences on collaborative recall. *Cognitive Development, 20*, 33-48. Doi:

10.1016/j.cogdev.2004.07.002 .

Leman, P. J., Ahmed, S., & Ozarow, L. (2005). Gender, gender relations, and the social dynamics of children's conversations. *Developmental Psychology*, *41*, 64-74. doi: 10.1037/0012-1649.41.1.64

Leman, P. J., Skipper, Y., Watling, D., & Rutland, A. (2016). Conceptual change in science is facilitated through peer collaboration for boys but not for girls. *Child Development, 87*, 176-183. doi: 10.1111/cdev.12481 .

Leman, P. J. & Tenenbaum, H. (2017). Communication in children's and adolescents' social groups, in A. Rutland, D. Nesdale, & C. S. Brown (Eds.), *The Wiley-Blackwell Handbook of Group Processes in Children and Adolescents*. Oxford: Wiley-Blackwell.

Leung, K. C. (2015). Preliminary empirical model of crucial determinants of best practice for peer tutoring on academic achievement. *Journal of Educational Psychology, 107*, 558-579. doi: 10.1037/a0037698

Light, P., & Littleton, K. (1994). Cognitive approaches to group work. In P. Kutnik & C. Rogers (Eds). *Groups in schools*, (pp. 87-103). London: Cassell.

Littleton, K., Light, P., Joiner, R., Messer, D., & Barnes, P. (1992). Pairing and gender effects on children's computer-based learning. *European Journal of Psychology of Education, 7*(4), 311-324. doi: 10.1007/bf03172896

Maitland, K. A., & Goldman, J. R. (1974). Moral judgment as a function of peer group interaction. *Journal of Personality and Social Psychology*, *30*(5), 699.

Messer, D. J., Joiner, R., Loveridge, N., Light, P., & Littleton, K. (1993). Influences on the effectiveness of peer interaction: Children's level of cognitive development and the relative ability of partners. *Social Development, 2*, 279-294. doi: 10.1111/j.1467-9507.1993.tb00018.x

Meudell, P. R., Hitch, G. J., & Boyle, M. (1995). Collaboration in recall: Do pairs of people cross-cue each other to produce new memories? *The Quarterly Journal of Experimental Psychology, 48*, 141-152. doi: 10.1080/14640749508401381

Murphy, N., & Messer, D. (2000). Differential benefits from scaffolding and children working alone. *Educational Psychology, 20*, 17-31. doi: 10.1080/014434100110353

Palincsar, A. S. (1998). Social constructivist perspectives on teaching and learning. *Annual Review of Psychology, 49*, 345-375.

Piaget, J., & Inhelder, B. (1956). The child's conception of space. New York: W. W. Norton.

Piaget, J. (1932). *The Moral Judgment of the Child*. London: Routledge & Kegan Paul.

Piaget, J. (1959). *The Language and Thought of the Child.* London: Routledge & Kegan Paul.

Piaget, J. (1985). *The equilibration of cognitive structures: The central problem of intellectual development*. Chicago: University of Chicago Press.

Psaltis, C., Duveen, G., & Perret-Clermont, A. N. (2009). The social and the psychological: Structure and context in intellectual development. *Human Development, 52*, 291-312. doi: 10.1159/000233261

Radziszewska, B., & Rogoff, B. (1988). Influence of adult and peer collaborators on children's planning skills. *Developmental Psychology, 24*, 840-848. Doi: 10.1037//0012-1649.24.6.840

Reznitskaya, A., Anderson, R. C., McNurlen, B., Nguyen-Jahiel, K., Archodidou, A., & Kim, S.-y. (2001). Influence of oral discussion on written argument. *Discourse Processes, 32*(2-3), 155-175. doi: 10.1080/0163853x.2001.9651596

Rohrbeck, C. A., Ginsburg-Block, M. D., Fantuzzo, J. W., & Miller, T. R. (2003). Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Educational Psychology, 95,* 240–257. doi: 10.1037/0022-0663.95.2.240

Rosenthal, R. (1991). Meta-analysis: a review. *Psychosomatic Medicine, 53, 247-271*. doi:

10.1097/00006842-199105000-00001

Roseth, C. J., Johnson, D. W., & Johnson, R. T. (2008). Promoting early adolescents' achievement and peer relationships: The effects of cooperative, competitive, and individualistic goal structures. *Psychological Bulletin, 134,* 223-246. doi: 10.1037/0033-2909.134.2.223

Russell, J., Mills, I., & Reiff-Musgrove, P. (1990). The role of symmetrical and asymmetrical social conflict in cognitive change. *Journal of Experimental Child Psychology, 49,* 58-78. doi: 10.1016/0022-0965(90)90049-e

Sills, J., Rowse, G., & Emerson, L. M. (2016). The role of collaboration in the cognitive development of young children: a systematic review. *Child: Care, Health and Development*, *42* 313-324. doi: 10.1111/cch.12330

Slavin, R. E. (1992). "When and why does cooperative learning increase achievement? Theoretical and empirical perspectives". In R. Hertz-Lazarowitz & N. Miller (Eds.), *Interaction in cooperative groups: The theoretical anatomy of group learning* (pp. 145–173). New York: Cambridge Univ. Press.

Strough, J., Berg, C. A., & Meegan, S. P. (2001). Friendship and gender differences in task and social interpretations of peer collaborative problem solving. *Social Development*, *10*, 1-22. Doi: 10.1111/1467-9507.00145

Tenenbaum, H. R., Ford, S., & Alkhedairy, B. (2011). Telling stories: Gender differences in peers' emotion talk and communication style. *British Journal of Developmental Psychology, 29*, 707-721. doi: 0.1348/2044-835x.002003

Tenenbaum, H. R., Leman, P. J., Aznar, A., & To, C. (2016). Studying children’s conversation. In J. Van Herwegen and J. Prior (Eds.), *Developmental Methods.* Oxford: Psychology Press.

Thanh, P. T. H., Gillies, R., & Renshaw, P. (2008). Cooperative learning (CL) and academic achievement of Asian students: A true story. *International education studies*, *1*, 82-89. doi: 10.5539/ies.v1n3p82

Thompson, R. B., & Thornton, B. (2014). Gender and theory of mind in preschoolers' group effort: Evidence for timing differences behind children's earliest social loafing. *The Journal of Social Psychology*, *154*, 475-479. DOI: 10.1080/00224545.2014.933763

Tolmie, A., Thomson, J. A., Foot, H. C., Whelan, K., Morrison, S., & McLaren, B. (2005). The effects of adult guidance and peer discussion on the development of children's representations: Evidence from the training of pedestrian skills. *British Journal of Psychology, 96*, 181-204. doi: 10.1348/000712604x15545

Topping, K., Buchs, C., Duran, D., & Van Keer, H. (2017). *Effective peer learning: From principles to practical implementation*. London: Routledge.

Tudge, J., & Rogoff, B. (1999). Peer influences on cognitive development: Piagetian and Vygotskian perspectives. In Peter Lloyd, Charles Fernyhough (Eds). *Lev Vygotsky: Critical Assessments*, (pp. 32-56). London: Routledge.

Underwood, J., Underwood, G., & Wood, D. (2000). When does gender matter?: Interactions during computer-based problem solving. *Learning and Instruction*, *10*, 447-462. doi: 10.1016/s0959-4752(00)00008-6

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Mental Process*. Cambridge, MA: Harvard University Press.

Vygotsky, L. S. (1981). The instrumental method in psychology. *The concept of activity in Soviet Psychology*, *137*, 143.

Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research, 13,* 21–39. doi: 10.1016/0883-0355(89)90014-1

Weinstein, B. D., & Bearison, D. J. (1985). Social interaction, social observation, and cognitive development in young children. *European Journal of Social Psychology, 15*, 333-343. doi: 10.1002/ejsp.2420150307

Wood, D., & Middleton, D. (1975). A study of assisted problem-solving. *British Journal of Psychology*, *66*, 181-191.

Table 1. Moderators

Moderators	List of Categories
Assessment Pattern	Pretest Posttest
	Posttest Only
Time of assessment	During
	Immediately after
	Delay
	No information
	Combined
Comparison	Individual children
	Adults
Same task	No
	Yes
	For 75%
	For 50%
Consensus	Asked to agree
	No instructions about consensus provided
Age	Young (up to and including age 10 years)
	Old (11 years to 18 years)
	Mixed (study includes both old and young children)
Gender of Pairs	Mixed
	Same
	Both mixed and same
	Unknown
Group Size	Two
	More than two
Area of Learning	Conservation
	Creativity
	Mathematical Reasoning

	Memory
	Moral
	General Reasoning
	Scientific reasoning
	Spatial Conservation
	Spatial Reasoning
Publication characteristic	Published
	Unpublished

Table 2. List of Studies

Author(s)	**Total *n***	***Hedge's g*** [LL, UL]	**Post-test delay**	Assessment pattern	**Same type of task**	**Adult Peer**	**Consensus**	**Age**	**Gender**	**Size**	**Area of Learning**
Ames & Murray (1982)	48	.71 [.12/1.31]	delay	Pre	Yes	individuals	Yes	Y	M	2	CN
Arterberry, Cain, & Chopko (2007)	192	.26 [-.04/.56]	during	Post	Yes	individuals	No	Y	S	2	GR
Asterhan, Schwartz, Cohen-Eliyahu (2014)	390	.03 [-.23/.29]	delay	Pre	Yes	individuals	No	O	NR	2	GR
Bearison, Magzamen, & Filardo (1986)	99	-.07 [-.48/.34]	delay	Pre	Yes	individuals	Yes	Y	S	2	SC
Blaye, Light, Joiner & Sheldon (1991)	38	.95 [.16/1.74]	delay	Post	Yes	individuals	Yes	O	S	2	SLR
Botvin & Murray (1975)	36	3.05 [2.1/4]	immed	Post	No	individuals	Yes	Y	NR	>2	CN
Chan (2001)	108	-.04 [-.41/.33]	immed	Pre	Yes	individuals	No	O	MS	2	SR
Damon & Killen (1982)	147	-.17 [.45/.80]	delay	Pre	Yes	mixed	Yes	Y	M	>2	ML

Diehl & Strobe (1987) study 1	42	-1.86 [-3.21/-.51]	immed	Post	Yes	nominals	No	O	S	>2	CY
Doise & Mugny (1979)	72	.42 [-.06/.9]	delay	Pre	Yes	individuals	Yes	Y	NR	2	SC
Doise, Murray & Perret-Clemont (1975) study 1	60	1.13 [.33/1.92]	during	Post	Yes	individuals	Yes	Y	S	2	SC
Doise, Murray & Perret-Clemont (1975) study 2	49	0.71 [.11/1.32]	Delay	Pre	No	individuals	Yes	Y	NR	>2	CN
Druyan (2001)	338	-.69 [-1.22/-.14]	Delay	Pre	Yes	both	Yes	Y	NR	2	SR
Duran & Gauvain (1993)	48	.54 [-.05/1.13]	immed	Pre	Yes	individuals	No	Y	S	2	SLR
Emler & Valiant (1982) study 1	80	.017 [-.42/.45]	Delay	Pre	Yes	individuals	Yes	Y	NR	2	SLR
Fawcett & Garton (2005)	100	.95 [.50/1.46]	Delay	Pre	Yes	individuals	No	Y	S	2	GR
Garton & Pratt (2001)	222	.02 [-.26/.31]	Delay	Pre	Yes	individuals	No	Y	S	2	GR
Gauvain & Rogoff (1989) study 1	48	.07 [-.73/.58]	during	Post	Yes	individuals	No	Y	S	2	SLR
Gauvain & Rogoff (1989) study 2			during	Post	Yes	mixed	No	Y	S	2	SLR

Gauvain & Rogoff (1989) study 2			immed	Post	Yes	mixed	No	Y	S	2	SLR
Golbeck (1998)	40	.41 [-.28/1.10]	immed	Pre	Yes	individuals	Yes	O	S	2	SC
Golbeck (1998)	45	-.43 [-1.07/.21]	immed	Pre	Yes	individuals	Yes	Y	S	2	SC
Gummerum, Leman, & Hollins (2013)	312	-.34 [-.97/.29]	during	Post	Yes	nominals	No	Y	S	>2	MY
Howe, McWilliam, & Cross (2005) study 2	138	.36 [.023/.69]	Delay	Pre	some	individuals	Yes	M	NR	>2	SR
Howe, Taylor, & Devines (2013)	139	.37 [-.27/.39]	Delay	Pre	some	individuals	No	M	NR	>2	SR
Howe, McWilliam, & Cross (2005) study 3	143	.21 [-.12/.53]	Delay	Pre	Yes	both	Yes	Y	NR	2	SR
Howe, Taylor, & Devines (2016)	168	.06 [-.45/.48]	Delay	Pre	some	mixed	No	Y	NR	2	SR
Hudgins (1960)	128	.28 [-.067/.63]	no info	Post	Yes	individuals	Yes	M	NR	>2	MR
Kirschner, Paas, & Kirschner (2009)	70	.43 [-.38/1.25]	no info	Post	Yes	individuals	No	O	NR	>2	SR

Kruger (1992)	72	.6 [.11/1.10]	immed	Pre	Yes	adult	Yes	Y	S	2	ML
Kuhn, Shaw, & Felton (1997)	49	.62 [-.13/1.37]	no info	Pre	No	individuals	Yes	O	S	2	GR
Leman and Oldham (2005)	96	-.35 [-.75/.047]	during	Post	Yes	nominals	No	Y	S	2	MY
Leman, Skipper, Watling, & Rutland (2016)	324	.38 [.16/.60]	Delay	Pre	No	individuals	No	Y	S	2	SR
Light & Glachan (1985) study 1	32	.90 [.21/1.59]	Delay	Pre	Yes	individuals	Yes	O	NR	2	GR
Light & Glachan (1985) study 1	32	1.27 [.50/2.04]	Delay	Pre	Yes	individuals	Yes	Y	NR	2	GR
Littleton, Light, Joiner, Messer, & Barnes (1992)	120	.46 [.00/.83]	Delay and during	Pre	Yes	individuals	No	O	MS	2	SLR
Lumpe & Staver (1995)	52	1.61 [.88/2.34]	Delay and during	Pre	Yes	individuals	Yes	O	NR	>2	SR
McCall (2017)	66	2.46 [1.81/3.12]	During	Post	Yes	individuals	Yes	O	NR	>2	SR
Maitland & Goldman (1974)	30	-.09 [-1.03, .84]	Delay	Pre	Yes	individuals	No	O	NR	>2	ML

Maitland & Goldman (1974)	30	.18 [-.76, 1.12]	Delay	Pre	Yes	individuals	Yes	O	NR	>2	ML
Manion & Alexander (1997)	90	.34 [-.072/.75]	immed	Pre	Yes	individuals	Yes	Y	NR	>2	MY
Messer, Joiner, Loveridge, Light, & Littleton (1993) study 1	40	-.33 [-1.07/.42]	immed	Pre	Yes	individuals	No	Y	S	2	SR
Meudell, Hitch, & Boyle (1995)	96	1.34 [.81/1.87]	during	Pre	Yes	individuals	Yes	O	NR	2	MY
Murphy & Messer (2000)	122	.39 [-.82/.03]	immed	Pre	Yes	both	Yes	Y	NR	2	SR
Perlmutter, Behrend, Kuo, & Muller (1989) study 3	42	.16 [-.49/.81]	Delay	Post	Yes	individuals	No	Y	MS	2	SR
Phelps & Damon (1989)	110	.52 [.12/.93]	no info	Pre	Yes	individuals	No	Y	S	2	MR
Psaltis & Duveen (2002)	213	.50 [.22/.78]	Delay	Pre	No	individuals	Yes	Y	MS	2	CN
Psaltis (2011)	266	.39 [-.04/.81]	Delay and immediate	Pre	No	individuals	Yes	Y	MS	2	SC
Radziszewska & Rogoff (1988)	32	-.66 [-1.35/.039]	immed	Post	Yes	adult	No	Y	S	2	SLR

Radziszewska & Rogoff (1991)	40	-1.18 [-1.84/-.52]	immed	Post	Yes	adult	No	Y	S	2	SLR
Roazzi & Bryant (1998)	78	1.55 [1.04/2.06]	Delay	Pre	Yes	individuals	No	Y	NR	>2	SR
Russell (1981)	161	.30 [-.082/.67]	during	Pre	Yes	individuals	Yes	Y	S	2	CN
Russell, Mills, & Reiff-Musgrove (1990) study 1	99	.93 [.29/1.58]	during	Pre	Yes	individuals	No	Y	S	2	CN
Russell, Mills, & Reiff-Musgrove (1990) study 2	64	-.06 [-.79/.66]	during	Pre	Yes	individuals	No	Y	S	2	CN
Russell, Mills, & Reiff-Musgrove (1990) study 3	64	2.20 [.64/3.77]	during	Pre	Yes	individuals	No	Y	S	2	CN
Samaha & DeLisi (2000)	88	.67 [-.92/2.57]	During and immediate	Pre	Yes	individuals	Yes	O	both	>2	GR
Schwarz (1995) study 1	28	2.67 [.99/4.35]	during	Post	Yes	individuals	Yes	O	S	2	SR
Schwarz (1995) study 2	91	.73 [.19/1.27]	during	Post	Yes	individuals	Yes	O	NR	2	SR

Schwarz (1995) study 3	40	1.43 [.50/2.35]	during	Post	Yes	individuals	Yes	O	NR	2	SR
Schwarz & Linchevski (2007)	60	.60 [.084/1.11]	Delay	Pre	No	individuals	Yes	O	NR	2	MR
Stacey (1992)	80	-.001 [-.44/.44]	during	Post	Yes	individuals	No	O	NR	>2	MR
Teasley (1995)	46	.63 [.13/1.13]	immed	Post	Yes	individuals	No	Y	S	2	SR
To (2015)	112	.15 [-.22/.52]	immed	Pre	Yes	individuals	No	O	MS	2	SR
Tolmie, Thomson, Foot, Whelan, Morrison, & Mclaren (2005) study 1	38	-.14 [-.71/.43]	Delay	Pre	some	mixed	No	Y	M	>2	GR
Torrance (1970)	39	.47 [-.11/1.04]	during	Post	yes	individuals	No	Y	NR	2	CY
Tudge (1989)	84	.08 [-.61/.77]	Delay	Pre	yes	individuals	Yes	Y	NR	2	SR
Tudge (1992)	148	-.10 [-.41/.22]	Delay	Pre	yes	individuals	Yes	Y	S	2	SR
Tudge & Winterhoff (1993)	79	.01 [-.54/.56]	Delay	Pre	yes	individuals	Yes	Y	S	2	SR

Tudge, Winterhoff, & Hogan (1996)	140	.36 [-.094/.82]	Delay	Pre	yes	individuals	Yes	Y	S	2	SR
Valiant, Glachan, & Emler (1982)	78	.93 [.38/1.48]	Delay	Pre	yes	individuals	Yes	Y	NR	2	SC
Weinstein & Bearison (1985)	80	1.66 [1.17/2.16]	Delay	Pre	yes	individuals	Yes	Y	S	2	CN
Williams & Tolmie (2000)	96	1.29 [.75/1.83]	Delay	Pre	yes	individuals	Yes	Y	M	>2	SR
Zawilinski (2012)	75	.28 [-.37/.91]	immed	Pre	no	individuals	No	Y	MS	2	GR

Note. For Age, Y = 4 to 10 years; O = 11 to 18 years; M = spans both age groups. For Area, CN = Conservation, CY= Creativity, G = General Reasoning, MR= Mathematical Reasoning, MY= Memory, ML = Moral, SCN = Spatial Conservation, SLR = Spatial Reasoning, SR= Scientific Reasoning. For Gender, MS = Mixed- and same-gender groups, S = Same-gender groups, M= Mixed-gender groups, NR= Not reported.

Table 3. Moderator Effects

	Hedge's g	UL/LL	*Z*	*k*	Heterogeneity Q	Comparison by Moderator *Q*
Assessment Pattern				2		.01
Posttest Only	.42	.06 / .78	2.26*	21	157.82***	
Pretest Posttest	.40	.27/ .53	6.17***	50	195.68***	
Time of Posttest						4.85
During	.68	.30 / 1.05	3.54***	16	100.78***	
Delay	.41	.25 / .58	4.68***	31	134.82***	
Immediately	.10	-.26 / .46	.53	15	86.36***	
Adult versus Child Comparison						12.40***
Children	.45	.32/.58	6.85***	68	37.93***	
Adults	-.30	-.70/.10	-1.48	10	295.91***	
S task Comparison						.80
No	.54	.25 / .83	3.66***	9	36.52***	
Yes	.39	.23 / .55	4.76***	57	300.57***	
Consensus						11.74**
Yes	.61	.42/.81	215.63***	39	6.12***	
No	.17	.02/.33	117.29***	33	2.14*	

Age						2.73
Older Age Group (11-18 years)	.62	.32 / .91	4.08***	20	109.10***	
Younger Age Group (4- 10 years)	.33	.18 /.49	4.18***	48	240.03***	
Gender of Peer Groups						1.82
Mixed	.70	.07/1.34	2.17*	5	19.16***	
S	.24	.05 / .44	2.43***	33	140.82***	
Group size						1.88
More than 2 children	.61	.25 / .97	3.29***	18	128.89***	
Two children	.34	.20 / .47	4.93***	53	207.08***	
Area of learning						11.05
Conservation	.98	.51/ 1.47	4.04***	9	52.58***	
Scientific Reasoning	.45	.22 / .69	3.73***	24	141.62***	
General Reasoning	.39	.12 / .65	2.83**	10	26.38**	
Spatial conservation	.37	.01/.72	2.04*	7	18.83**	
Mathematical Reasoning	.34	.09 / .58	2.66**	4	4.19	
Moral	.31	-.14 / .67	1.73	3	2.38	
Memory	.25	-.49 / .98	.65	4	28.25***	
Spatial reasoning	-.03	-.47/ .41	-.14	8	28.88***	
Creativity	-.61	-2.89 / 1.66	-.53	2	9.60*	

Note. * $p < .05$, ** $p < .01$; *** $p < .001$

Figure 1. *PRISMA checklist for systematic review of the literature*

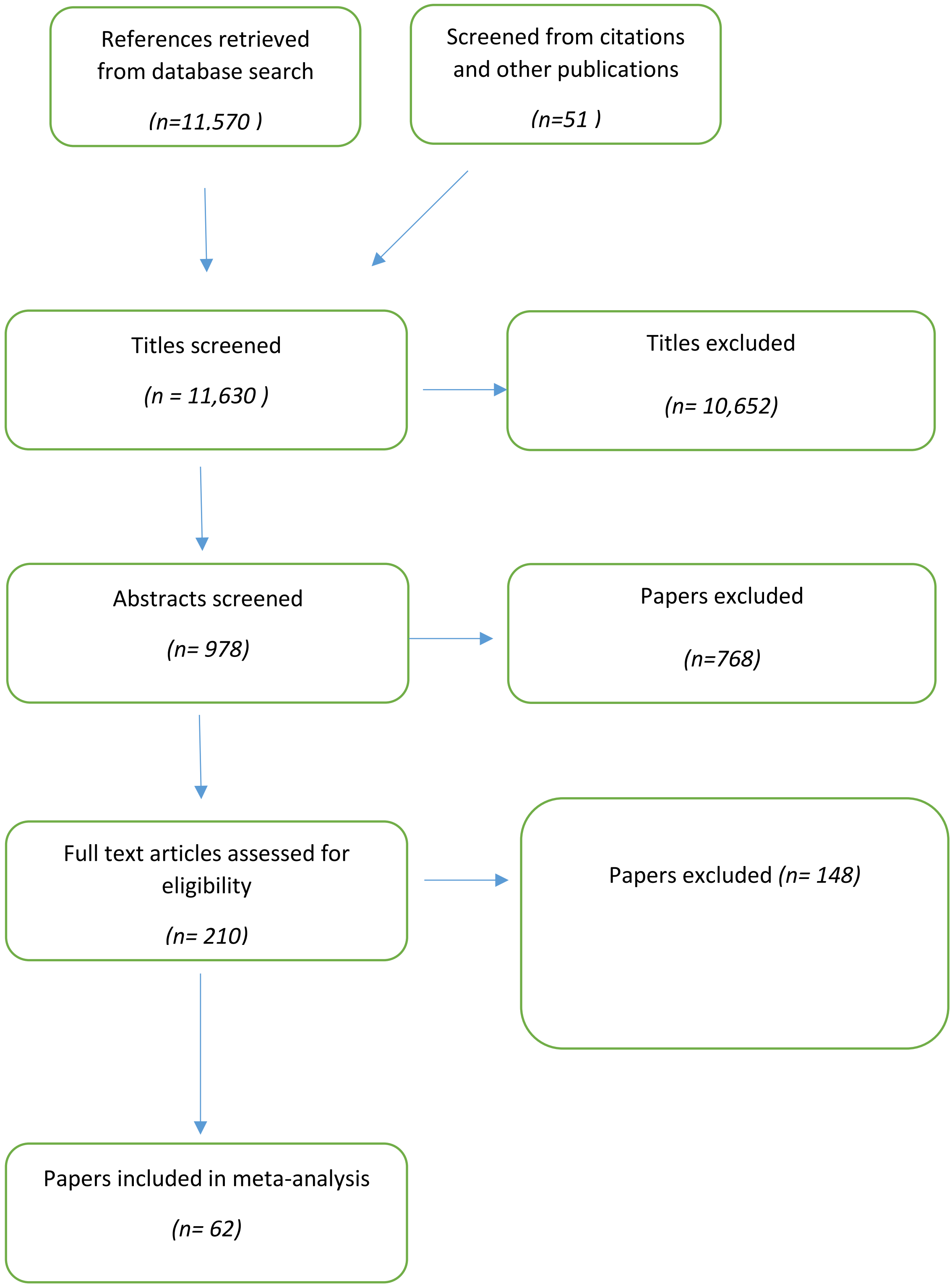

Figure 2. Funnel plot